

**National Assessment Governing Board**

**Meeting of March 2-3, 2001**

**Coral Gables, Florida**

**OFFICIAL SUMMARY OF BOARD ACTIONS**

Complete Transcript Available

**Present**

**NAGB Board Members**

Mark Musick, Chairman  
Michael Nettles, Vice-Chairman  
Moses Barnes  
Melanie Campbell  
Wilmer Cody  
Daniel Domenech  
Edward Donley  
Thomas Fisher  
Edward Haertel  
Juanita Haugen  
Nancy Kopp  
Ronnie Musgrove  
Debra Paulson  
Jo Ann Pottorff  
Diane Ravitch  
John Stevens  
Migdania Vega  
Deborah Voltz  
Michael Ward  
Marilyn Whirry  
Dennie Palmer Wolf

**NAGB Board Members Absent**

Lourdes Sheehan  
Adam Urbanski

**National Assessment Governing Board (NAGB) Staff**

Roy Truby, Executive Director  
Sharif Shakrani, Deputy Executive Director

Tessa Campbell  
James Carlson  
Mary Crovo  
Ray Fields  
Lawrence Feinberg  
Munira Mwalimu

**National Center for Education Statistics (NCES) Staff**

Gary Phillips, Acting Commissioner  
Peggy Carr, Associate Commissioner  
Steve Gorman  
Suzanne Triplett

**American Institutes for Research (AIR) Staff**

George Bohrnstedt  
Steve Ferrara  
Fran Stancavage

**Council of Chief State School Officers (CCSSO) Staff**

Wayne Martin  
John Olson

**Educational Testing Service (ETS) Staff**

Jay Campbell  
Jeff Haberstroh  
Archie Lapointe  
Steve Lazer  
John Mazzeo  
Paul Ramsey

**Westat Staff**

Dianne Walsh

**Attending Speakers**

Katherine Blasick, Broward County, Florida, Public Schools  
Michael Casserly, Council of the Great City Schools  
Martha Coxe, Florida Department of Education  
Sharon Lewis, Council of the Great City Schools  
Harold O'Neil, University of Southern California/CRESST

## **Others**

Lois Peak, National Education Goals Panel  
Donna Simmons, Office of Governor Ronnie Musgrove, Mississippi  
Larry Snowwhite, McA Enterprises  
Brad Thayer, National Computer Systems

## **Call to Order**

The March 2, 2001 meeting of the National Assessment Governing Board was called to order at 8:30 a.m.

## **School Visit Report**

Migdania Vega provided a briefing on the Board members' March 1 visit to Coral Way Bilingual Elementary School. Ms. Vega reported that the school is comprised of 1,463 enrolled students and has a special bilingual program with instruction provided in two languages. The school has set high standards and the school's philosophy is that children, regardless of their socioeconomic backgrounds, will perform well when standards are set high. Coral Way Bilingual Elementary School is a Title 1 school, with 71 percent of students qualifying for free and reduced lunch. Seventy percent of the students are immigrants. Further, many of the students come from low social economic homes. Despite these factors, the school has received an "A plus" grade in the state of Florida for the year 2001, which qualified the school to receive a grant of \$132,000. Ms. Vega remarked that this accomplishment evidences the fact that students from low socioeconomic backgrounds can in fact perform well if high standards are set and if one works hard with the students. Ms. Vega described the core knowledge curriculum, and explained features of the international studies program.

Michael Nettles thanked Ms. Vega for the hospitality provided to Board members during the school visit.

## **Review and Approval of Agenda**

Mr. Musick reviewed the meeting agenda. A motion for Board approval of the agenda was made, seconded, and carried unanimously.

## **New Board Member: Introduction and Oath of Office**

Mr. Musick introduced new Board member Ronnie Musgrove, Governor of Mississippi and administered the Oath of Office to swear in the Governor as a Board member. Governor Musgrove remarked that he grew up in rural Mississippi and was a first generation college graduate in his family, earning undergraduate and law degrees at the University of Mississippi. Governor Musgrove had served two terms as a Senator and was Chairman of the Education

Committee, and during his second term significant legislation was passed on education initiatives. Governor Musgrove credited his first grade teacher for her dedication and for motivating him to perform well. The Governor commended President Bush for placing utmost importance on education.

### **Executive Director's Report**

Roy Truby provided a report on the following items:

- The Board had hired two new staff members—Jim Carlson, Director of Psychometrics and Munira Mwalimu, Operations Officer;
- New Board Member orientation sessions were held with Governor Musgrove, Roy Nageak, and Lourdes Sheehan;
- Board member terms of office had been extended by one year to enable members to serve four year terms;
- Mark Musick and Roy Truby had met informally with Education Secretary Rod Paige and had extended an invitation for the Secretary to attend the May Board meeting scheduled to take place in Annapolis, Maryland;
- Mr. Musick and Mr. Truby had also met with the President's transition team and groups that had expressed concern on the National Assessment of Educational Progress (NAEP) becoming akin to a national curriculum. Mr. Musick and Mr. Truby held discussions related to their concerns; and
- The National Education Goals Panel Task Force had recently convened and put forth very bold proposals that were supportive of NAEP and that were designed to secure NAEP participation and reduce the testing burden. Mr. Truby highlighted some of these proposals and noted that this type of support from the Goals Panel during this time of reauthorization comes at an opportune time. It will also assist the Board in responding to the President's "No Child Left Behind" proposal.

### **Welcoming Remarks and Comments**

Martha Coxe, Deputy Commissioner for the Florida Department of Education, welcomed Board members to Florida on behalf of the Governor of Florida Jeb Bush, Lieutenant Governor Frank Brogan, and Commissioner of Education Charlie Christ.

Ms. Coxe described Florida's ambitious education program known as the *A Plus Program*, which is a system of assessing schools on an annual basis. The state has a rigorous standards driven curriculum and schools are held to a certain standard. School performance against state standards is measured through statewide assessments, and grades are assigned. Rewards are

made to schools that earn a grade of A, or that make exceptional improvement of at least one letter grade. Intense assistance is offered to schools that do not meet the standard. While the measurement concept is simple, it is not as simple to implement. Refinements to the program continue to be made. Ms. Coxe illustrated the point that in the U.S. Marines, as in the school system, uniform standards are set and while different persons learn at different levels, it is important that instruction be provided for those who need it in order to complete the programs of instruction. Ms. Coxe indicated that students have a variety of skill sets, some are disadvantaged, and all learn at different paces. It is the job of the education system to ensure that students leave school being able to compete in the world and to be successful; failure is not an option.

Florida fully supports the President's "No Child Left Behind" initiative as well as the NAEP program. In response to concerns about too much testing, Ms. Coxe stated that assessments provide knowledge on student performance that ultimately assists in helping children learn more.

Ms. Coxe commended Board member Thomas Fisher's contribution to the development of Florida's *A Plus Program* and for the states' success in education assessment and measurement.

### **NAEP Update**

Gary Phillips, Acting Commissioner, National Center of Education Statistics (NCES), provided the following updates:

- Reports to be released in 2001 include the following:

#### **NAEP Releases:** (at press conferences)

- Late March or early April: The NAEP Reading Report Card;
- July: The Math Report Card;
- September: The Science Report Card;

#### **Third International Mathematics and Science Study (TIMSS) Releases:**

- March 15: International Civics Assessment;
- April 4: Benchmarking study (released by Boston College);
- April 27: A U.S. report on the International Civics Assessment (released at the Education Writers' meeting in Phoenix, Arizona);
- July: The third release of the early childhood longitudinal study in kindergarten (which will be followed up with a birth study);
- November: TIMSS video study;
- 2001: Item information on the 1988 NAEP Civics Assessment; and
- Latter part of 2001: Report on the Program for Indicators of Student Achievement (PISA), which is a new international, study involving 29 countries, including the U.S. The survey will be repeated every three years in science, math, and reading, focusing on a different topic each year.

Mr. Phillips reported that new surveys that are forthcoming include the adult literacy survey

to be conducted in 2002 and possibly another TIMSS study in 2003.

- NCES is currently in the process of revising and updating its methodology standards, some of which relate to NAEP. This process is expected to be completed by the end of the year.

Mr. Phillips noted that while NAEP received an increase in funding for fiscal year 2001 from \$68 million to \$80 million, the fiscal year 2002 budget will require \$93 million in order to complete all the work that NCES is being asked to do. If this amount is not funded, priorities will have to be established.

### **Update on the Administration's Education Proposal**

Mr. Musick remarked that the update and discussion on the Administration's education proposal is intended to focus the Board's attention on how President Bush's education proposal relates to or affects the NAEP program. Although the proposal is in its formative stages, Board members need to understand the specific components of the proposal vis-a-vis the NAEP program.

Mr. Truby reported that the Senate education bill was expected to be on the Senate floor the week of March 5. The Board needs to start thinking of NAEP related matters such as what "confirming" really means, what is the gap, and when does one know that it is closed? While there were many unanswered questions about the President's proposal, NAGB needs to start thinking about these issues. Mr. Truby stated that Mr. Shakrani of NAGB and Ms. Carr of NCES will share information that may assist in responding to the education proposal, and Mr. Fields of NAGB will provide illustrative information by comparing a sample state's data to NAEP's data.

Sharif Shakrani summarized President Bush's "No Child Left Behind" initiative as it relates to accountability. The initiative aims to improve the performance of all students, especially disadvantaged students and closing the achievement gap between the high achieving and the low achieving students. To achieve this goal, the President is proposing that states have annual testing in reading and mathematics in grades three through eight, and in conjunction, annual state NAEP testing in reading and mathematics in grades four and eight. NAEP is to be used to "confirm" state results. Mr. Shakrani commented that the concept of "confirming" is still being studied. He highlighted the following technical issues pertaining to the proposal:

- (1) The NAEP program will continue and will not change as a result of the new initiative since the President's proposal is above and beyond what is currently in place.
- (2) The current framework developed by NAEP, the sampling technique, current testing window, and reporting by achievement levels will continue.
- (3) As it is currently structured, NAEP is intended to provide an overall picture of how students are doing, and is not intended to measure the gap. As a result, the sample size may need to be reconsidered to produce valid and reliable results, especially if NAEP is used for confirming purposes.
- (4) NAEP mathematics and reading items must be sufficient to measure the total domain of the framework as adopted by the Board.

- (5) The items must be secure to produce reliable estimates of what students can do at the state level in grades four and eight.
- (6) Since there is likely to be more focus on NAEP, it will be very helpful to release more items so that educators can know what is being measured, with examples of what is being tested.
- (7) Since testing will take place every year, field testing should be integrated with the operational assessment.

Peggy Carr commented on the following technical aspects of the proposal, which are consistent with the Board's initiatives on state participation and reporting timelines:

- (1) A major technical issue to be resolved is how to do the sampling for the national and the state NAEP. One possibility is to combine state and national standards so that the data can be used together. For this to work, all 56 jurisdictions would have to participate. Standardization of the national and state components would also be necessary. This approach would greatly reduce the burden on school administrators; there would be uniform treatment on exclusion; and using only one contractor for the data collection would reduce the risk of test items being leaked, stolen, or copied.
- (2) It will be necessary to develop a timeline for an expedited analysis of the data. This will require a better way to get the data from the schools to the schooling centers in the least amount of time. Using the Internet is a possibility, as long as security precautions are used.
- (3) Ms. Carr pointed out that another issue to consider is replacement of released items. She noted that open-ended items are difficult to get through the approval process expeditiously, and as a result, there could be fewer open-ended items in the future.

Mr. Shakrani stated that the sample size affects confirmation. He highlighted the desired sample size specifications, and emphasized the importance of communicating to states how the data are reported and how disadvantaged is defined in NAEP. These are important steps prior to the final confirmation stage.

There are additional research issues to be considered such as accommodation (since there are different rules of exclusion used by NAEP and the states); participation; motivation; and the increased testing burden. Edward Haertel suggested adding another research issue—examining the variant testing windows and exploring the possibility of having schools used in two-year cycles, so that 50% of the sample would be the same in two successive years, significantly increasing the accuracy of trends.

Diane Ravitch asked what percentage of test items are released after the NAEP assessment. Mr. Shakrani responded that items are never reused after they have been released, and that between 25% and 40% are released when the assessment is completed. Ms. Ravitch remarked that NAEP's confirming role will raise the stakes and impact student motivation. She therefore suggested further research in this area.

Ray Fields presented a simulation of NAEP data vis-à-vis a sample state's data to illustrate what a confirmation might look like, and to illustrate issues and concerns that may be raised in the process.

Mr. Fields stated that there are many questions to consider such as definitions of “achievement gap,” “confirmation,” and the criteria or standards to be used in making those determinations. It is also unclear what year the starting point/baseline will be.

Mr. Fields emphasized that the simulation was a work in progress and is for illustrative purposes only. He went on to illustrate the data for "State A" versus the NAEP data. Mr. Fields pointed out that the states have 50 unique state testing programs, different standards, different definitions of disadvantaged students, and of adequate yearly progress. Further, currently NAEP does not have a way to confirm the adequate yearly progress of a state using the state's definition.

Board members then engaged in a question and answer session on the simulation. Topics discussed included:

- How the achievement gap would be defined;
- The definition of disadvantaged students;
- Different definitions of performance levels among states vis-a-vis NAEP, which will affect the reporting of accurate results;
- The definition of adequate yearly progress and who sets the definition;
- The possibility of fewer open-ended test items in the future and resulting impact on measuring more complex skills; and
- Differences in what NAEP and state assessments measure.

Ms. Ravitch suggested that, given the new kinds of responsibilities and challenges that NAGB is facing, a committee of the Board should be established to take another look at NAGB's frameworks to ensure that the frameworks which were written 10 or 11 years ago are, in fact, appropriate for the new responsibilities.

Mr. Fields thanked NCES and ETS staff for their assistance with the presentation. Mr. Cody thanked Mr. Shakrani and Mr. Fields for their research work on these complex issues, and he emphasized the need for all the issues to be resolved very carefully. Mr. Truby thanked Peggy Carr and the contractors for their work on the model.

### **Meeting Recess**

The first session of the March 2, 2001 Board meeting recessed at 10:35 a.m. to permit various committees to meet.

### **Meeting Reconvened**

The second session of the March 2, 2001 Board meeting reconvened at 1:45 p.m.



## **Update on 2004 Mathematics Framework Project**

Wayne Martin, the Council of Chief State School Officers (CCSSO), provided an update on the 2004 Mathematics Framework Project. As a result of requests made by various states, CCSSO has chosen to maintain the existing framework with some modifications so that the short-term trend lines from 1994, upon which the state NAEP is based, will be carried forward. Another project responsibility is to design a small-scale study of advanced topics at the 8<sup>th</sup> grade level to ascertain the curriculum at this grade level.

The Council of Chief State School Officers has partnered with the Council for Basic Education (CBE) and the Association of State Supervisors for Mathematics (ASSM) to undertake this project. Mr. Martin briefed the Board on the staff members assigned to work on the project that is scheduled to be completed within 14 months.

A planning document was submitted by CCSSO to the Board in October. A background Issues Paper that discusses the status of curriculum and standards in math since 1990, is in the final draft stage. The object of the project is to develop recommendations for an updated mathematics framework and for the item and test specifications. The test developer will use this blueprint, which is being developed through a consensus process. The final Mathematics Assessment Framework is scheduled to be completed by July 20. The test and item specifications will be completed by September 14, and the final report will be submitted to NAGB by November 30.

John Olson, CCSSO, reported that a Steering Committee, a Planning Committee, and a Technical Advisory Panel have been established to guide the project. He described the composition of the two Committees and of the Panel and their areas of responsibilities. Mr. Olson also reported that the ASSM is donating its assistance as it can provide a link to the states. The Association is convening ten regional meetings to gather input from state supervisors on state math standards, assessments, uses of NAEP, reporting data, and how NAEP ties into state approaches. This information is being collected at both local and state levels. Data collected by ASSM will be provided to CCSSO to inform the project.

Mr. Truby commented that when he and Mr. Musick met with various groups at the White House, they shared information with the groups on how the NAEP framework development process is a bottom-up and not top-down process. The grassroots process was looked upon very favorably, allaying concerns that the framework would be an imposition of a national curriculum.

Mr. Martin indicated that CCSSO would have a progress report at the May 2001 Board meeting. Mary Crovo reported that draft documents on the project would be presented for Board review at the August Board meeting. Board action on the framework and specifications is scheduled to take place at the November 2001 meeting.

Wilmer Cody commented that the special probes provide an opportunity on a fairly economical basis to not only get the measures of what students know and can do, but gather other information about why they know those things and what is going on in the instruction program.

He felt that there is far more that NAGB should be exploring, and urged richer data collection in smaller segments of the population, as opposed to that done on a large-scale by NAEP.

### **Trial Urban Assessment: A Panel Discussion**

Michael Nettles introduced Michael Casserly, Executive Director of the Council of Great City Schools (CGCS), Katherine Blasick, Executive Director of Research, Evaluation, and Student Assessment in Broward County, Florida, and Sharon Lewis, Director of Research, CGCS.

Mr. Nettles reminded the Board that the proposal for a Trial Urban Assessment that was presented to the Board at its November 2000 and March 2001 Board meetings is continuing to be explored. The Board found the proposal to be intriguing and was working with NCES staff to explore its feasibility. In addition to the ten districts that were initially proposed for the assessment, the Board is exploring the idea of including five additional districts. NAGB is moving in the direction of establishing an Advisory Committee to make recommendations to the Board about specific policies and practices governing the administration of NAEP in the districts.

Michael Casserly described the composition of the Council of the Great City Schools (CGCS) and membership criteria. He stated that CGCS is a coalition of the nation's largest urban public school systems across the country, with a membership of approximately 58 major city school systems. The Council serves as a national voice for urban education, and has an active set of supports for building capacity and assistance to local school systems across the country. The Council's work is divided into four major categories: improving student achievement and closing the academic achievement gaps in the cities; improving leadership and governance and management of urban public school systems; improving opportunity, financing, and funding of urban schools; and providing professional development for all of the constituents inside of the urban community. The Council also has an active and aggressive research effort and is supported and funded directly by the individual school systems.

Mr. Casserly then summarized the Council's proposal to design and conduct a trial NAEP assessment of large, urban public school systems across the country. He explained that the proposal was initiated for three reasons:

- (1) The Council is committed to setting high academic standards and improving the quality of urban education in the U.S. Currently, there is no way of measuring progress on those standards. The Council is unable to answer the question, that if progress is being made, who is making the fastest progress and who is closing the achievement gaps the most? Another question to be answered is, of all the school districts making really fast progress, why are some school districts making headway and others are not?
- (2) The Council needs to collect additional information on what is effective in closing the academic achievement gaps in the districts. Data needs to be collected across state lines, city by city.
- (3) More comparable data is needed on communities similar to urban school districts to enable comparisons, such as comparing data from New Orleans to data from St. Louis.

Mr. Casserly stated that since the proposal for the assessment was first presented to the Board in November 2000, there have been some developments. The Board's Executive Committee has approved a resolution to proceed with the proposed assessment. Further, CGCS has presented the proposal to President Bush's transition team. Mr. Casserly was hopeful that the detailed budget to be submitted to Congress in the coming weeks will include funding for this initiative.

Mr. Casserly introduced Katherine Blasick, Director of Research, Evaluation, and Testing for the Broward County (Florida) Public Schools, which is the fifth largest school system in the nation.

Ms. Blasick described Florida's state assessment—the FCAT, which is an annual statewide assessment at multiple grade levels in mathematics, reading, and writing, with social studies and science to be included in the near future. In addition to a criterion reference type of assessment based on the state's standards, Florida also administers a reference assessment to students in grades 3 through 10. Broward County has also implemented the SAT 9 to 1<sup>st</sup> and 2<sup>nd</sup> grade students, because the county cannot wait until the end of 3<sup>rd</sup> grade to get that kind of information in order to gauge where its students are performing in comparison to the nation. In addition, 10<sup>th</sup> graders this year took the state sponsored PSAT on a voluntary basis, with a 40-45% participation rate.

The FCAT is a high stakes test as results are used for making promotion and retention decisions. Multiple pieces of information are being looked at. Students who are not demonstrating what they need to know in order to be successful at the next grade level are retained, and a summer program of instruction is provided. The criteria part of the assessment is also used as a graduation requirement. The FCAT results are used for grading schools through letter grades—A through F.

Ms. Blasick noted that Broward County also looked at what was happening nationally and recognized that not only does the county want to know about the proportion of students that hit a certain standard, but also recognized that with the large number of students (approximately 250,000) coming into its school system, it also must look at gain, because some students are not coming in at the same level as other students.

It is important to have information on individual students over time. It was Ms. Blasick's viewpoint that the highest stakes on any assessment is that people make judgments about the schools based on the assessment results. The key is for the assessments to help identify the strengths and weaknesses of individual students as well as in the overall curriculum, and to institute changes so as to improve student performance and close the achievement gap.

Ms. Blasick pointed out that an upcoming issue for the state is that the FCAT data are going to be used for teacher evaluation. It is important to be able to compare districts and schools so that positive experiences and gains can be replicated or variables that can help increase student achievement can be understood.

Ms. Blasick reported that the state data and individual district data are reported publicly on the states' web site and through the media. Test results are electronically transferred to the county

so that they are available to the teachers quickly. At any point in time a teacher can pull data for their class, disaggregate their class data, and look at similar students across the school. Ms. Blasick stressed the importance of receiving the results back in time for the classroom teachers or district administrators to make decisions to improve what they are doing as soon as possible. The state is also spending a lot of time on professional development to help teachers understand the information and to make good decisions about instruction based on data.

Sharon Lewis of the Council of Great City Schools provided a summary of the Atlanta assessment system, on behalf of Atlanta Superintendent Beverly Hall who was unable to make her presentation to the Board in person.

Atlanta administers a kindergarten assessment test, which is a behavioral checklist; a writing assessment in grades 3, 5, and 8; as well as the Stanford 9 complete edition in the same grades. The criterion reference test, developed by the State of Georgia, will be administered for the first time in April of 2001 to students at grades 4, 6, and 8 in language arts, reading, and mathematics. In 2002, it will be expanded to include more grades. Georgia also administers a high school graduation test starting at grade 11 and it also participates in the state sponsored PSAT. There are also a number of local assessments, including an algebra I exam, the scholastic reading test, the scholastic mathematics test, and an end-of-year course testing in mathematics.

Georgia participates in the NAEP assessment, and Atlanta has, for many years, participated in the Georgia sample. However, the Atlanta Public Schools do not really use NAEP results, because they do not receive individual student, school or system score reports.

Mr. Casserly asked Ms. Blasick how data from the Trial Urban Assessment conducted by NAEP would assist Broward County and how the issue of test burden would be addressed. Ms. Blasick responded that NAEP data would allow comparison of state vis-à-vis national student performance and results could be used to continually strengthen curriculum, identify areas of need, and allow a focus in those areas. She also hoped that a Trial NAEP might provide additional key information to look at the national data to see if there is something different within the urban districts. If information is made available by district, it will enable districts to ascertain where they are meeting with greater success and share the keys to success. On the issue of test burden, Ms. Blasick stated that she did not envision much additional burden since her district already participates in NAEP.

Thomas Fisher wondered what Ms. Blasick would say to her staff and students so that they would approach the NAEP activity with the same gusto with which they approach the FCAT. Ms. Blasick replied that she would have to convince her teachers that the test results would have value, such as helping to make curriculum changes and decisions on programming. The teachers would then have to present it to the students as being important. Another issue to consider is timing of the NAEP assessments in relationship to other state assessments.

Board members then engaged in a question and answer session on the proposed Trial Urban Assessment. Issues addressed included the choice of subjects assessed and participation of other large school districts that are not city based such as counties or large suburban school systems.

Roy Truby commented that showing an overall direction and being able to confirm a direction is going to be easier technically and conceptually than looking at the gap. To look at the issues raised in a systematic way, Chairman Musick has appointed an Advisory Panel to be chaired by Michael Nettles, with members to be decided. Mr. Truby asked if there was some way, as part of the Trial Urban Assessment, to better examine the gap issues, how to better display the gap, and to somehow dovetail this proposal into a model to take a better look at what NAGB has been asked to do or may be asked to do with the Bush initiative. Mr. Casserly acknowledged that the gap issue needs to be addressed further and that it is not currently possible to make gap closing comparisons across states for the major cities. It would be a major benefit to have a Trial Urban NAEP Assessment that specifically addresses the issues.

Mr. Casserly expressed concern that the use of NAEP as a confirmation would serve as a veto on whether or not a state's progress has been rewarded. For example, if the state assessment showed progress by the state but the NAEP assessment did not show progress, whatever bonus, reward or recognition the state might receive towards progress would then not go forward.

Jo Ann Pottorff asked Mr. Casserly to explain the difference in memberships between CGCS and the Council for Urban Boards of Education. Mr. Casserly explained that the Council for Urban Boards of Education is an affiliate of the National School Boards Association, and its membership consists of individual school board members, not cities, as with CGCS. He stated that CGCS attracts the largest school districts, while the Council for Urban Boards of Education attracts the mid- and moderately-sized cities across the nation. With CGCS's membership, the administration, the superintendent, and the Board of Education work closely to address issues of urban education.

Diane Ravitch observed that many districts do not want to know their results, because they are afraid that they will be embarrassing. She therefore congratulated the Council for its courage in wanting to know the results, and its willingness to live with the results and use them.

Ms. Wolf remarked that two years ago, Mr. Casserly had raised the question about differentiating below basic, and now the Board has a proposal in front of it, not only to differentiate basic but also the proficient range. Ms. Wolf expressed the possibility that this trial could be an arena in which the Board could look at how that differentiation works out. Mr. Casserly acknowledged that this would be very constructive, as CGCS would be interested in that kind of differentiation across all of the levels, particularly in below basic, because a large share of the students are below basic, and there is no other way of measuring their progress up the ladder.

Mr. Cody asked if CGCS will be able to pay attention to trend data when it provides the reports to cities on their state assessment systems. Mr. Casserly assured him that there will be trend data, and each city will be benchmarked against the state trend, subject-by-subject, and grade-by-grade.

## **Closed Session**

### **NAEP 2000 Reading Report Card: Fourth Grade**

In accordance with the provisions of exemption (9)(B) of Section 552b(c) of Title 5 U.S.C., the full Board met in partially closed session on March 2, 2001 from 12:15 p.m. to 1:35 p.m. to receive a briefing from Gary Phillips, Acting Commissioner of the National Center for Education Statistics on the National Assessment of Educational Progress (NAEP) 2000 Reading Report Card: 4<sup>th</sup> grade national results.

Mr. Phillips described the features and contents of the NAEP 2000 national reading assessment. Results were presented in the following categories:

- Average scale scores for the nation;
- Scale score percentiles;
- Achievement level results;
- Results by race/ethnicity and gender;
- Trends in racial/ethnic/gender gaps;
- Average scale scores by type of school;
- Pages read daily for school and homework;
- Time spent on homework;
- Reading for fun;
- Number of different types of reading materials in the home;
- Time spent watching TV daily;
- Percentage of students by race/ethnicity;
- Students with disabilities; and
- Students with Limited English Proficiency

## **Open Session**

### **Review NAEP Long-Term Trend Assessment Items**

Peggy Carr described the history and nature of the long-term trend assessments in reading, science, math, and writing. The assessments consist of frameworks that were developed in the 1960s and 1980s and represent the more basic skill knowledge base requirements that were important for students to know then. The newer and more contemporary frameworks are more advanced and reflect more advanced level skills. The items that appear on the long-term trend assessment are the original items. At one point, there was an attempt to update and change the items, but after the 1986 reading anomaly, NCES put those items back on the assessment.

Ms. Carr summarized the differences between the main NAEP assessment and the long-term trend assessment:

- (1) An important feature of the long-term trend assessment is in the nature of how the items are actually administered to the students. They are administered in balanced and complete

blocks; i.e., the math and science are administered together, and the reading and writing are administered together, which is different from how the main assessment is administered.

- (2) The main NAEP assessment and more contemporary frameworks are at grades 4, 8, and 12, as opposed to the model grades of 4, 8, and 11 that are represented in the long-term trends.
- (3) The long-term trend takes about 45 minutes to administer, and has very few open-ended items on the assessment, in sharp contrast to the more contemporary assessments which have more stated response questions (short and long) and open-ended items.
- (4) Student and school questionnaires are included on the long-term trend assessments, but there is no teacher questionnaire.
- (5) Proficiency scores are reported on a scale from 0 to 500, with a cross-grade scaling approach, which means there is a hierarchical structure to the numerical reporting of the data by age level.
- (6) The major subgroups include whites, blacks, and Hispanics.
- (7) Private schools are included in the assessment, but unlike the main NAEP assessment, there is no attempt to do an oversampling for private schools.

The assessment was last given in 1999 and is due to be administered again in 2003. Ms. Carr noted that studies show that what appears on the long-term trend represents those skills and knowledge that are still important to our children, and that they are really under the more contemporary framework in the assessments that represent them.

Ms. Ravitch commented that the long-term trend assessment covers only very basic skills, and educators feel they have progressed past that to much more advanced student skills. Ms. Ravitch questioned why the long-term trends are so flat, instead of reflecting a steady increase in the performance if they are really assessing the lower level skills. Ms. Carr responded that in the lower levels, such as level 150, representing the more basic skills than the overall assessment itself, evidences an increase in the number of students achieving those levels all the time, so there has been some progress in basic skills.

### **Closed Session**

#### **Review of NAEP Long-Term Trend Assessment Items**

In accordance with the provisions of exemption (9)(B) of Section 552b(c) of Title 5 U.S.C., the full Board met in partially closed session on March 2, 2001 from 3:30 p.m. to 5:00 p.m. to review secure items on the NAEP Long-Term Trend Assessment. Items were reviewed in the subjects of Mathematics, Science, and Reading in three separate breakout sessions.

### **Open Session**

#### **Review of NAEP Long-Term Trend Assessment Items**

The Board reconvened in open session to continue its discussion on item review of the long-term

trend assessment.

Thomas Fisher noted that in his review of the long-term trend items, he came across several instances in which items might be considered to be somewhat out of date. There were also instances where the student performance might have been actually lowered because they might not have fully understood the instructions due to the page layout. He expressed disappointment that the long-term trend reveals very little improvement. Mr. Fisher reported that his group had discussed whether or not maintaining a long-term trend of this nature is worth the bother; Board members were more concerned than pleased with the item review.

Marilyn Whirry stated that her group enjoyed seeing the trend line, even if it was not positive, especially in reading and mathematics, where there is not the archaic quality of some of the questions, as seen in the science items. The consensus of her group was to continue the development of the trend line, at least in reading and mathematics.

Mr. Musick and Mr. Haertel pointed out that the item level results are much less stable than the assessment level results.

Mr. Cody commented that many of the science items seem outdated, and warned that it would be wrong to conclude that students do not know any more science today (since the trend is flat) because the scope of what is to be taught has probably changed. He felt that the science trend data is misleading, and that NAGB ought to explore the possibility that the lack of progress or change in trend is a function of change in the instructional program. He also suggested the possibility of dropping the trend and starting all over again.

Juanita Haugen stated that she felt the same way about the reading and math items. She noted that classroom instruction has changed over the years, and as a result, she felt that students would view the questions as irrelevant with very little meaning to them. Debra Paulson also noted the change in instruction, but stated that the students should still be able to respond to most of the questions. She acknowledged that some of the questions are irrelevant and some might not be understood by the students, because they are shown in a different way than the way students are taught.

Edward Haertel commented on the diminishing alignment with science, and stated that if the current NAEP assessments were given to students back in 1977 and 1984, they would probably do much worse on the current assessments than students are doing today on the old assessments, which he viewed as encouraging. He pointed out that if students can still do those things now just as well as they could back then, and they are also doing more now, that in itself is an important message. Mr. Cody pointed out that the students have to demonstrate that they are doing more now. Mr. Haertel also noted that the P values were useful.

Ms. Whirry, responding to Ms. Haugen's comments, noted that the reading test is very similar to the 1984 test and that some of the reading passages were very good. What the long-term trend assessment is missing is the richness of the NAEP main test, because there are no open-ended questions. Her group felt differently from Ms. Haugen's group and liked many of the questions.



Mr. Haertel suggested studies that could be done with these data, including item parameter drift, which is a kind of modeling—an IRT-based procedure, to see whether individual items are getting harder or easier relative to the backdrop of the other items over time. He felt that these data are an ideal place to do that. He noted that it had been done for the writing trend, but that it had been discontinued.

Mr. Haertel commented that someone in his group expressed concern on different trends in P values on the constructed response/multiple choice items in writing. He speculated that one reason may be that, especially at the fourth grade level, students are more likely to be asked to write their responses now than they used to be. He expressed an interest in going back to the long-term trend data and disaggregating that data to see whether or not there are differential trends for those two item formats.

Mr. Nettles stated that the Board continues to express a desire to maintain trends. However, using the same trend items over and over, say for the next 30 years, will be very outdated. He suggested that the Board explore alternate ways of maintaining trends in the future, without having the identical items over and over again. Mr. Nettles questioned whether there is some kind of way to link new items to the old assessments to maintain the trend line. Mr. Musick responded that this may be possible in math but he was not sure about reading and science.

Ms. Ravitch noted that some ideas being taught, such as percentages, are still necessary and are not antiquated. If they ever become outdated, they could be dropped at that point, but that point has not yet arrived. She stated that most of the questions she saw are still basic computational questions that children need for advanced problem solving, without which they will not be able to do the problem solving.

Ms. Haugen acknowledged that the question is not whether the skills are skills students ought to know and be able to do, but rather it is how the questions are presented that is outdated. She wondered if students could be told that the test was part of a long-term study so that they could understand the context in which the questions were presented.

Mr. Cody stated that he did not have a problem with the math and reading items, as with the science items. He suggested conducting an initial step analysis to compare what percentage of the main NAEP assessment is covered by the NAEP long-term trend in its alignment study. If the conclusion is that the long-term trend in science covers just a certain portion of what NAGB is trying to determine in the main science, then the question is how it is reported so that it is more accurate. There has been no change in the science trend, but there has been a major difference in science that is being measured.

Mr. Musick noted that 80% of the science domain was being measured 20 years ago, but today only 40% of the science domain is being measured, this fact that needs to be communicated.

Nancy Kopp reiterated concern that the science items do not measure what students know. She stated that the trend appears as if it is a smaller and smaller part of the domain. Something ought

to be done about it, because the results are misleading.

Mr. Stevens expressed appreciation of the opportunity to review the items as he now has a better understanding of what the NAEP Long Term Trend Assessment is all about.

Mary Crovo thanked ETS staff for their assistance in enabling Board review of the items. She stated that NAGB has some work underway to look at the alignment between main NAEP and the Long-Term Trend in those subjects. Results should be available at the May 2001 Board meeting.

### **Meeting Adjourned**

The March 2, 2001 session of the Board meeting adjourned at 5:35 p.m.

### **Meeting Reconvened**

The March 3, 2001 session of the Board meeting reconvened at 8:35 a.m.

### **Student Motivation and NAEP**

Harold O'Neil, Professor at the University of Southern California, provided a briefing on student motivation and NAEP.

He began by defining motivation as an internal state that arouses, directs and maintains behavior, which is usually contrasted to cognition. He added that one way to think about this is that cognition is thinking, motivation is feeling.

Mr. O'Neil pointed out that the TIMSS data show that 4th graders are above the international mean in math; 8th graders are at the mean; and 12th graders are far below the mean. One explanation for this is the fact that students tend to try less hard as they go through schooling and students in other countries maintain their motivation or are increasingly motivated as they go through school. Performance may be underestimated on a low-stakes test like NAEP, which brings about the question of whether NAEP findings accurately represent what students know and can do.

What is usually true of poor performance are cognitive issues such as students do not know enough, the academic engagement is low, there is a lack of teacher preparation, or a lack of standards. If this is true, that leads to a path of implementing cognitive fixes. To the degree that there are motivational influences, it leads down a different path to motivational fixes.

Mr. O'Neil theorized that students get more sophisticated as they get older and they decide to budget the resource they have, which is effort. For the high school senior going into the world of work or on to post-secondary education, tests like NAEP are low stakes. The cross-cultural data show that the students from Asian countries are doing better than we are, partly due to the fact that they are trying hard across the board. He pointed out that, if an Asian parent is asked why

their child is doing well in school, they tend to respond with “My kid works hard,” but if a U.S. parent is asked the same question, they tend to answer, “My kid is smart.” It is difficult to influence a student being smart, but it is fairly easy to influence a student to work harder.

To the degree that motivation is important, the desire is to increase student effort and self-efficacy. Another issue in the motivational area is that increasing the stakes for a proportion of the students also increases stress. Increased stress for some of those students results in increased anxiety, which, in turn, depresses performance. This is not true for all students, but perhaps for the top 20 percent of the distribution. Mr. O’Neil pointed out that another option to increase motivation is to provide incentives.

Mr. O’Neil mentioned a theory in the motivational arena that was adapted from a value and expectancy model. The expectancy asked the question “Can I do this task?” and the value question asks “Is it important to me?” Most theorists feel that this model is multiplicative, so that if one is zero, the performance will be low. Additional variables include gender and ethnicity. Another concept is that of trait state conception. Traits are pre-dispositioned, enduring characteristics in an individual, such as intelligence, achievement, or personality. They can be changed, but they are changed with difficulty.

The three major motivational influences on NAEP achievement at the time the test is being given consist of two positive influences—state self-efficacy (a student’s confidence of how well he is going to do on that task) and state effort (how much effort he puts into it)—and one negative, which is state worry. The two positive influences are self-efficacy and effort and there is one big negative, the worry business. Mr. O’Neil provided two examples:

- (1) If a female were taking a math test, there would be higher levels of trait worry than males, which would then raise the state worry, in turn reducing test performance for those students who were high in trait test anxiety.
- (2) Conversely, a student high in trait self-efficacy would demonstrate increased effort, which would produce better performance.

To increase motivation at test time, there are some things that can be done:

- Financial incentives (money for each correct item).
- Performance goal (competition).
- Mastery goal (personal accomplishment).
- Teacher-oriented goal (competition).
- Certificate of achievement incentive.
- Try your best (NAEP instructions).

Mr. O’Neil cited an NCES study performed in the early 1990s that looked at rewarding students for correct NAEP math items, using 8<sup>th</sup> and 12<sup>th</sup> graders. The study found that the financial incentives did not work with the 12<sup>th</sup> graders, but did work with the 8<sup>th</sup> graders. Unfortunately, there is not enough data available to understand why the incentives did not work with the 12<sup>th</sup> graders. He stressed the need for a research and development program to study these issues,

which involves three critical policy questions:

- (1) Are changes to research and development investments needed? There is very little research that explicitly addresses motivational issues at test time for tests such as NAEP. What is needed is a focused research and development program rather than the occasional study driven by individual investigators' intellectual interests.
- (2) How much research and development money is being spent on these issues? There is little to none being spent in industry and defense, and the same applies to the civilian sector.
- (3) How should we change our existing research and development investments to get answers to motivation and testing issues? During what time frame? What is the research and development road map? What is the cost? Who is in charge?

Questions that demand research and development answers include:

- (1) Can interventions be crafted at test time to increase state effort and state self-efficacy and reduce state worry and thus improve NAEP performance?
- (2) Do motivational processes and outcomes during NAEP testing vary as a function of gender or ethnicity?
- (3) How do possible new computer-based item formats (simulations, search tasks, knowledge maps) for NAEP relate to motivational processes and outcomes?
- (4) What is the impact of motivation on NAEP if NAEP becomes a high-stakes measuring rod?
  - Several states have made their testing programs more high stakes, and thus could be expected to demonstrate differences in motivation compared to low-stakes NAEP.
  - Would such high-stakes NAEP increase student worry; and if worry reduces NAEP performance, then would such high-stakes testing underestimate achievement?
  - Anxiety or worry is higher for Latino students and female students in general. Will these students be particularly hurt by high-stakes testing?

Mr. O'Neil reiterated that an entity needs to be established for identifying and prioritizing research across agencies. He estimated an annual cost of \$10-\$12 million to answer the questions he had put forth.

Board members then engaged in a question and answer session on motivation issues. Mr. Musick closed the discussion by asking Board members to consider further work or action the Board needed to pursue on the issue of motivation.

### **NAEP Interactive**

Steve Lazer, ETS, noted that two of the Board's redesign goals were to make the NAEP data more available to broader audiences, and to facilitate faster and more flexible reporting of that

data. Prior efforts to disseminate data have focused on three distribution methods:

- (1) Printed reports;
- (2) Restricted use data tapes; and
- (3) Data almanacs.

Mr. Lazer described the evolution of the data almanacs to the information that is currently available on the NAEP web site.

He then provided a detailed demonstration of NAEP Interactive, an online system under development to provide customized search results and greater flexibility of data. One of the reasons for developing this system is to help meet the six-month reporting goal. Mr. Lazer noted that states want shorter and snappier state reports with only five or six key variables, but every state wants five or six *different* key variables. The new system will enable states to write their own reports. The system will be introduced with the release of the math results, and will eventually include everything—the long-term trend study and the main assessments, and it will go back to the beginning of the current framework.

#### **Board Actions and Committee Reports**

The Board heard reports from its standing Committees. The reports from the Committee meetings, as revised and adopted after Board discussion, are appended to these minutes.

#### **Meeting Adjourned**

The March 3, 2001 session of the Board meeting adjourned at 12:05 p.m.

I certify the accuracy of these minutes.

---

Mark Musick, Chair

---

Date